Секция «История»

Информационная система «газета «Пермские губернские ведомости» как средство для сохранения и анализа объявлений.

Кузнецова Елена Александровна
Студент
Пермский государственный университет им. А.М. Горького,
историко-политологический факультет, Пермь, Россия
E-mail: opyat-26@yandex.ru

Объявления в периодических изданиях являются специфическим источником информации. Специфика состоит, во-первых, в том, что объявления массовый источник. Во-вторых, объявления содержат «непрямую» информацию. Исходя из этого, можно говорить о том, что возможно и необходимо использование компьютерных технологий для сохранения и анализа объявлений.

Компьютерные технологии создают возможность для более результативного анализа объявлений. Применение компьютерных технологий для анализа объявлений позволяет получить как количественные, так и качественные их характеристики.

В современной исторической науке актуальной становится задача более глубокого исследования уже известных (изученных) источников. Применение компьютеризированного анализа дает возможность извлечь скрытую информацию, содержащуюся в объявлениях.

В настоящей работе рассматривается комплекс объявлений, которые публиковались в блоке рекламных объявлений газеты «Пермские губернские ведомости». Рекламные объявления располагались на первой и последней странице «Ведомостей». На первой странице блок объявлений мог располагаться под заголовком на две трети или на всю страницу. На последней странице объявления занимали нижнюю половину или весь объем листа. Объявления могли располагаться и на $2-4$ страницах, в этом случае они занимали крайнюю полосу (от корешка).

Исследование осуществляется на основе информационной системы «Газета «Пермские губернские ведомости», созданной для сохранения электронных версий одноименной газеты и их анализа.

Основной единицей ввода информации в систему является публикация. Применительно к нашей задаче – это отдельное объявление. Система создает возможность сохранения и представления в полнотекстовом (распознанное содержание) и в графическом формате.

В процессе исследования решаются следующие задачи:

1. Перевод источника в машиночитаемый формат, ввод объявлений в систему.

2. Анализ комплекса объявлений как исторического источника, с помощью запросов к базам данных информационной системы.

Поскольку для сохранения и визуализации каждое объявление вводится в систему отдельно, блок объявлений, расположенный в электронных экземплярах Пермских ведомостей, с помощью программы ABBYY FineReader разрезается на отдельные статьи (объявления), которые параллельно сохраняются в двух форматах «PDF» и Microsoft Word.

Затем осуществляется ввод каждого объявления в информационную систему в виде распознанного текста и изображения. При вводе в систему каждому объявлению присваиваются следующие атрибуты: тип публикации, раздел, заголовок, примечание, дата номера (номер, дата), распознанное содержание, имидж статьи. Атрибут «Раздел» позволяет отделить объявления от других типов статей номера газеты. В поле «Тип публикации» введены типы объявлений, размещаемых в «Пермских губернских ведомостях» - это поможет провести группировку и количественный анализ блока объявлений. Поле «Заголовок» содержит запись, отражающую основной смысл объявления. «Дата номера» позволяет привязать текст объявления к номеру и странице газеты. Информация в поле «Распознанное содержание» позволяет производить контекстный поиск и формирование текстовых выборок для последующего контент анализ текста объявлений. Поле «Имидж статьи» необходимо, поскольку исследование ориентировано на источник, для наиболее точного сохранения его облика и содержания. Для атрибуции используются справочные таблицы системы, куда вводятся соответствующие стандартными значениями.

В зависимости от того, каково содержание объявления, выделены следующие группы: Медицина; Культура, искусство; Инфор.сообщения; Производство; Прочие услуги; «Требуется»; «Желаю получить место»; Продажа; Купля; Сдается; Сниму; Утеряно; Найдено; Образование; Соц.помощь, меценатство; Другое. Их имена и определены в качестве значений для поля «Тип публикации». Система позволяет производить классификацию объявлений путем группировки их на основании указанных значений.

На примере блока объявлений, расположенного в газете «Пермские губернские ведомости», № 7 от 10 января 1909 г., рассмотрим возможности информационной системы.

Можно осуществить классификацию объявлений в соответствии с атрибутами, назначенными каждому из них в поле «Тип публикации».

На основе результатов, полученных при классификации можно выяснить количественное соотношение различных типов объявлений. Так, выбрав нужный номер и дату, предлагаем системе найти все объявления, привязанные к этой дате. Получаем результат – 41 объявление располагались на первой и последней страницах номера.

Благодаря произведенной классификации объявлений, мы можем выяснить количественное соотношение разных типов объявлений в номере. Выбрав раздел «реклама» и нужный нам тип публикации (например «Продажа») узнаем количество представленных объявлений. Таким образом можно осуществить запросы и по другим значениям поля «Тип публикации». Итак, выясняем, что в данном номере представлены объявления следующих категорий: «Продажа» - 13, «Желаю получить место» - 4, «Требуется» - 2, Инфор.сообщения - 1, Медицина – 6, Образование – 3, Сдается – 6, Утеряно – 1, Культура, искусство – 3, Прочие услуги – 2. Таким образом, введя объявления за определенный период, мы сможем увидеть процентное соотношение различных сфер жизни города, представленных в объявлениях.

Полученные за определенный период результаты можно интерпретировать как характеризующие соотношение различных сфер жизни города, определенных социальных групп или видов деятельности и потребностей населения.

Система позволяет увидеть и то, как часто публиковались те или иные повторяющиеся объявления. Это позволяет делать выводы о том, насколько стабильны были потребности и интересы, предложение товаров и услуг в рассматриваемом социуме. Ре-

зультаты анализа объявлений позволяют интерпретировать и с точки зрения степени развития тех или иных сторон социально-экономической и культурной жизни города или иных населенных пунктов губернии.

Таким образом, использование компьютерных технологий для сохранения и анализа объявлений периодической печати является необходимым процессом. Это позволяет извлекать из источника скрытую информацию, осуществлять качественный и количественный анализ информации источника и получать более обоснованные результаты.